

# 연구단보: 얼굴 표정을 이용한 비접촉식 거짓말 탐지 시스템 개발 및 유의시점 분석

남보름 · 김주영<sup>1</sup> · 김명수<sup>1</sup> · 최진경 · 조백환<sup>2</sup> · 김인영<sup>1\*</sup>

한양대학교 일반대학원 융합전자공학부, <sup>1</sup>한양대학교 의생명공학전문대학원 생체의공학과,  
<sup>2</sup>삼성서울병원 AI 연구센터

## Technical Note

### Development of Non-contact Deception Detection using Facial Expression and Analysis of Significant Time Period

Bo Rum Nam, Joo Young Kim<sup>1</sup>, Myeong Su Kim<sup>1</sup>, Jin Kyoung Choi,  
Cho Baek Hwan<sup>2</sup>, and In Young Kim<sup>1\*</sup>

*Department of Electronic Engineering, Hanyang University, Seoul 04763, Korea*

*<sup>1</sup>Department of Biomedical Engineering, Hanyang University, Seoul 04763, Korea*

*<sup>2</sup>Medical AI Research Center, Samsung Medical Center, Seoul 06351, Korea*

*\*E-mail: iykim@hanyang.ac.kr*

(Received August 04, 2021; Revised August 07, 2021; Accepted August 30, 2021)

**요약:** 본 연구는 기존의 접촉식 거짓말 탐지의 단점을 해결하기 위하여 비접촉식 거짓말 탐지 기법을 개발하고자 얼굴 표정 변화를 촬영하고 분석하였다. 거짓말 탐지를 위한 영상 데이터 얻고자 모의범죄 실험을 진행하였으며, 긴장정점검사(POT)를 이용하여 데이터를 수집하였다. 거짓말을 할 때의 표정 분석을 위해 합성곱 신경망(CNN) 계열의 stacked hourglass 네트워크 4층을 쌓은 모델을 이용하여 얼굴에서 특징점을 추출하였다. 영상에서 프레임 단위로 추출한 66개의 특징점을 시간 순으로 배치하여 장단기메모리(LSTM) 모델을 이용하여 거짓말 영상과 진실 영상을 분류하는 모델을 개발하였으며, 모델 학습을 진행하고 이를 토대로 평가 데이터를 통한 분류 정확도는 85.19%이었다. 학습이 완료된 모델의 결과를 분석하기 위하여 LSTM 활성화 분석을 진행하였다. 개발한 방법을 이용하여 사람의 얼굴에서의 표정 변화를 이용한 거짓말 탐지의 가능성을 확인할 수 있었으며, 비접촉식 거짓말 탐지에 중요한 시점을 도출하여 추후 비접촉식 거짓말 탐지의 가능성을 확인할 수 있었다.

**Abstract:** In this study, facial expression changes were photographed and analyzed to develop a non-contact deception detection technique to solve the shortcomings of the existing contact deception detection. A simulated crime experiment was conducted to obtain video data for deception detection, and the data was collected using a POT (Peak of Tension Test). In order to analyze facial expression when lying, face landmarks were extracted from the face by using a model with four layers of stacked hourglass networks, a kind of CNN (Convolutional Neural Network). The deception detection was performed using stacked 66 face landmarks extracted in the frames from video in chronological order by the LSTM (Long Short Term Memory) model for truth/deception video classification, and the result using test set was 85.19%. LSTM activity analysis was performed to analyze the results of the model on which training was completed. Using the developed method, the possibility of deception detection using changes in facial expressions was confirmed, and important time for non-contact deception detection were identified.

**Key Words:** Forensic investigation, Deep-learning, Deception detection, Face landmark, Convolutional neural network, Long short term memory

기존의 거짓말 탐지기(Polygraph)는 거짓말을 할 때 나타나는 자율신경계의 변화에 따라 호흡, 맥박, 혈압, 피부저항 등을 감지하여 거짓말을 판별하게 된다. 이와 같은 다양한 생체신호를 측정하기 위해서 복잡한 사용절차를 거치게 되며 측정을 위해 부착한 다양한 센서들은 검사 대상자의 불필요한 긴장감을 일으켜 유도하는 거짓말 반응이 나오지 않아 검사에 차질을 일으킬 수 있다. 또한 부착식 센서들로 인하여 움직임이 제한되고 측정된 생체신호 해석에 검사관의 주관적인 판단에 따른 오류가 발생할 수 있다는 단점이 존재한다. 따라서 기존 거짓말 탐지기의 단점을 보완하기 위해 비접촉식 생체정보를 이용한 거짓말 탐지 시스템의 개발이 필요하다.

비접촉식 생체 정보 중, 본 연구에서는 얼굴 표정을 이용한 거짓말 탐지 알고리즘을 제시하고자 한다. 얼굴표정 변화는 얼굴의 눈썹, 눈, 눈동자, 미간, 코, 입술, 볼 등에 존재하는 안면 근육을 통해 나타난다.<sup>1</sup> 안면 근육으로 나타나는 다양한 표정들은 거짓말의 단서가 될 수 있는데 그 예로 매우 짧은 시간 안에 나타나는 미세표정 변화<sup>2</sup> 와 불안으로 인한 행동학적 지표,<sup>3</sup> 그리고 거짓말로 인한 감정 상태를 숨기기 위한 표정들이 있다.<sup>4</sup>

비접촉식 거짓말 탐지를 위해 모의실험을 진행한 연구<sup>5</sup>가 존재하지만 영상에서의 얼굴 표정 추출을 자동화하기 위하여 피험자의 얼굴을 움직이지 못하게 고정시켜 놓는 방법을 사용하였다. 따라서 이 방법은 기존의 접촉식 거짓말 탐지에서의 단점이었던 움직임을 제한한다는 점을 해결할 수 없었다. 그 외에도 사람이 직접 영상을 보고 특정 얼굴표정 존재 유무를 이용하는 방법<sup>6</sup>이 존재한다. 그러나 사람이 직접 영상을 보고 얼굴 표정을 확인하는 방식은 표정을 확인하는 과정에서 주관적인 판단에 의한 오류가 발생할 수 있다.

따라서 본 연구에서 비접촉식 거짓말 탐지를 위하여 제안하는 방법은 다음과 같다. 1) 거짓말 할 때의 얼굴 움직임을 자동으로 추출하여 영상만으로 피험자가 거짓으로 대답할 때의 반응과 진실로 대답 할 때의 반응을 분류하는 신경망 모델을 활용하여 거짓말과 진실의 분류 알고리즘을 개발할 가능성을 확인하고, 2) 개발한 알고리즘을 통해 그 결과를 분석하여

어느 시점이 가장 효율 적으로 거짓말을 판별할 수 있는지를 알아보고자 한다.

데이터 수집을 위해 신체 건강한 성인 27명을 대상으로 모의 범죄 실험을 진행하였다. 실험 절차는 다음과 같다. 피험자는 아무도 없는 실험실에 세 종류의 상품권 중 한 가지를 골라 본인이 원하는 곳에 숨긴다. 그 후 피험자에게 거짓말 탐지 검사 기법으로 사용되는 긴장 정점검사(POT: Peak of Tension Test)<sup>7</sup>를 실시한다. POT 질문은 모의 범죄에 사용된 3개의 상품권 이미지를 보여주며 숨긴 카드가 해당하는 상품권이 맞는지 묻는 내용으로 구성되었다. 이때 피험자는 모든 질문의 대답을 ‘아니오’로 동일하게 한다. 또한 분석하고자 하는 3개 key question의 앞과 뒤에 padding question으로 실험에 사용되지 않은 네 종류의 상품권을 숨겼는지 질문하였다. 따라서 key question은 7개의 질문 중 2, 4, 6번째에 시행된다. Padding question인 1, 3, 5, 7번째 질문은 실 분석에서 제외하였다. 피험자는 고정된 위치의 의자에 앉아 의자 중심으로부터 50 cm, 바닥으로부터 70 cm 거리에 있는 모니터를 통해 상품권 이미지를 확인할 수 있으며 실험진행자의 질문과 동시에 모니터에 해당질문에 일치하는 이미지가 제공된다.

질문이 끝난 후 피험자가 ‘아니오’라는 대답을 한 후에도 질문에 해당하는 이미지는 모니터에 유지된다. 피험자의 ‘아니오’ 대답이 끝난 후 10초 뒤에 다음 질문을 진행하였다. 총 3번을 반복하여 6개의 진실 대답과 3개의 거짓 대답을 취득하였다. 실험을 진행하는 동안 앞서 설명된 고정된 위치의 모니터 상단 중앙에 설치된 웹캠을 통하여 피험자의 얼굴을 640×480의 해상도로 녹화하게 되며 초당 프레임수를 30으로 고정시켜 놓았다. 표정을 이용한 거짓말 탐지가 본 연구의 목적이었으므로, ‘아니오’ 대답 이후에 발생한 ‘Fight or flight’ 반응으로 인한 감정 상태 변화에 따른 표정을 이용하기 위하여 ‘아니오’ 대답 시작부터 그 이후를 분석하였다. 피험자에게 마이크를 설치하여 영상과 동시에 음성 또한 수집하였으며, ‘아니오’라고 대답하는 시간이 평균 0.5초였으므로 모든 데이터 길이는 대답이 시작된 시점부터 대답 이후 반응이 포함되기 충분한 5.7초로 맞추었다. 일정한 길이

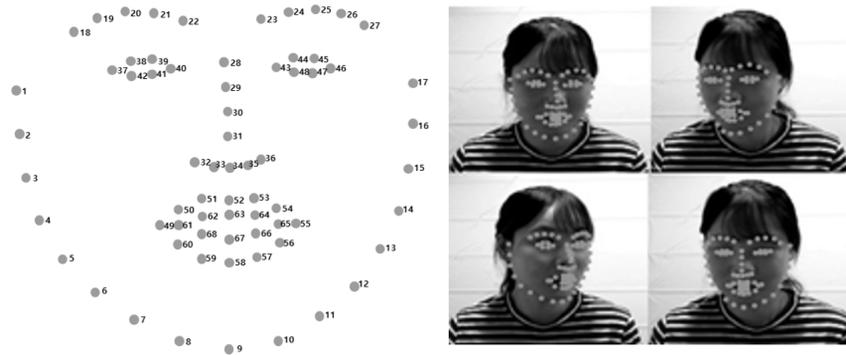


Fig. 1. 68 Face landmarks (Left) and face landmarks from stacked hourglass network applied to our data (Right).

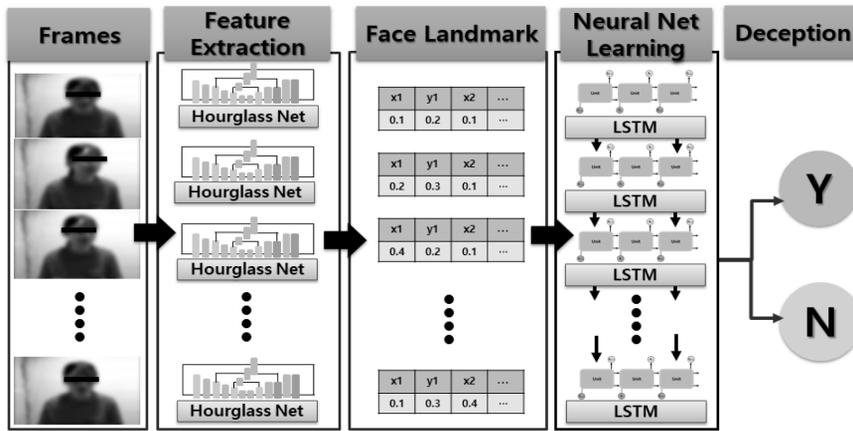


Fig. 2. Schematic diagram of the entire study method.

로 전처리 된 영상의 각 프레임에서 얼굴 특징점을 추출하기 위하여 CNN의 일종인 stacked hourglass 네트워크를 4층 쌓은 모델을 사용하였다.<sup>8</sup> Stacked hourglass 네트워크는 얼굴 특징점을 추출하는 모델 중 가장 뛰어난 성능을 보이는 네트워크이며, 추출한 얼굴 특징점과 실제 3D 카메라를 이용하여 추적한 얼굴특징점과의 오차율은 0.4% 내외이다. 추출된 68개의 얼굴 특징점(Face landmarks)은 코, 눈, 입, 눈썹, 턱의 x, y 좌표(Fig. 1)로 표현된다.

얼굴 특징점 데이터를 정규화하기 위하여 피험자 별로 눈과 눈 사이 거리를 이용하여 정규화를 진행하였다. 정규화에 사용된 값은 피험자의 오른쪽, 왼쪽 눈 동공 좌표의 유클리디안 거리를 이용하였다.<sup>9</sup> 딥러닝 모델 학습의 입력 값은 실험 절차 중 표정을 기록한 얼굴 특징점의 이전 프레임 값과 변화된 값의 차이를 계산한 값을 사용하였다. 계산된 차이 값은 68개의 얼굴 특징점(x, y) 좌표 136개의 x, y값의 이전

프레임 값과 변화된 값의 차이로 이미지에서 각 점의 x축 및 y축 변화 값을 의미한다. 또한 68개의 얼굴 특징점 중 얼굴 외곽 선 상부(점 1, 점 17)는 머리카락에 의해 오차가 컸기 때문에 최종 입력 데이터에서는 배제하였다. 영상에서의 각 프레임마다 66개의 얼굴 특징점을 시간 순으로 정리하여 최종 데이터로 사용하였다. 27 명에게서 얻어진 9개의 데이터를 24명의 train set(모델 학습용)과 3명의 test set(모델 성능 평가용)으로 무작위로 나누어 학습에 이용하였다.

거짓말 여부를 판별하기 위한 분류기로 딥러닝 신경망을 Fig. 2와 같이 이용하였다. 사용한 신경망의 종류는 시간 처리에 유리한 LSTM (Long Short Term Memory)을 사용하였다. 얼굴 특징점 66개를 입력 값으로 결정한 후, hold-out validation을 통해 모델을 학습하고 결과를 확인하였다. 그 결과 train set에 대한 정확도는 87.04%이었으며, test set에 대한 정확도

(accuracy)는 85.19%, F1-score는 79.65%이었다. 이는 진실/거짓 데이터 개수의 비율이 2:1이었을 때, 그 기회 수준에 해당하는 정확도가 66%이므로, 기회 수준보다 높은 정확도를 나타내며 딥러닝 신경망이 진실과 거짓의 각 특징에 따라 학습이 된 것을 확인할 수 있다. Grid search를 통해 결정된 LSTM은 1개의 layer와 70개의 hidden unit, 3개의 cell을 갖도록 설계하였으며 미니 배치 사이즈는 9, learning rate는 0.005로 1000 epoch 동안 학습 진행하였다.

모델의 학습 효율을 높이기 위하여 L2 regularizer를 사용하였으며 L2 regularizer의 lambda 값을 0.00001로 설정하였다. 거짓말과 진실을 영상으로 분류할 때 영상에서 모든 시점이 동일하게 중요한 것이 아니기 때문에 가장 중요한 시점을 계산하고자 하였다. LSTM이 영상 속의 표정 데이터를 순서대로 연산하여 거짓말 탐지를 진행하는 알고리즘이기 때문에, LSTM이 영상 중 어느 시점에 가장 집중하여 진실/거짓 여부를 판별하였는지를 확인하기 위해서 LSTM의 활성도를 분석하였다. LSTM cell activity는 하나의 동영상마다 시점별로 계산되게 되며, 그 동영상에 참인지 거짓인지를 판단할 때 중요한 시점에서 그 값이 커지게 된다. 따라서 LSTM cell activity가 클수록 LSTM 모델이 표정데이터를 시간 순서대로 분석하여 진실/거짓 여부를 판별할 때 집중한 시간대임을 이용하고자 하였고, 활성도가 큰 LSTM cell과 매칭 되는

영상 시점이 진실/거짓 여부 판별에 가장 중요한 시점이라고 판단하였다.<sup>10</sup> 각 영상에서 활성도가 큰 10개의 cell이 집중하고 있는 시간대의 값을 확인하였다.

Fig. 3에서 확인할 수 있듯이, 영상 데이터가 진실인 경우 전반적인 시간대에서 LSTM cell이 활성화되어, 영상이 진실/거짓임을 구분할 때 신경망이 가장 집중한 시점은 각각의 데이터마다 달랐음을 확인할 수 있었다. 영상 데이터가 거짓인 경우, 피험자의 대답이 30 프레임에서 45 프레임에 이루어질 때 그 이후에 LSTM cell이 활성화 되는 것을 확인할 수 있었다. Fig. 4와 Table 1에서 확인할 수 있듯이, 가장 유의한 시점은 대답 후 1초에서 1.5초 사이며 그 다음으로는 대답 직후부터 대답 후 0.5초 사이였다. 대답 전, 1초부터 대답 직전까지의 LSTM cell 활성화 도수는 3으로, 약 0.03의 상대도수를 가지나, 대답 직후부터 대답 후 1초의 LSTM cell 활성화 도수는 24로 약 0.27의 상대도수 값을 가지므로 거짓말 판별 시 대답 전보다는 대답 후의 표정 변화가 더 유의하게 작용하였음을 알 수 있었다. 또한 대답 직후부터 1.5초까지의 도수를 합하면 53으로, 이는 대답 직후부터 1.5초까지의 표정변화가 진실과 거짓말 구분에 다른 시점보다 더 중요하게 작용할 가능성이 크다는 것을 의미한다.

그러나 현재의 연구는 27명의 피험자에 대한 분석을 진행한 것이기 때문에 연구의 신뢰성을 위해 더

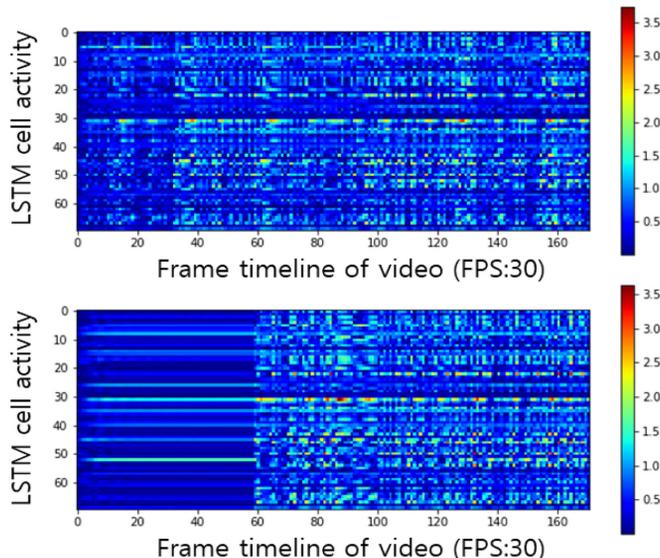
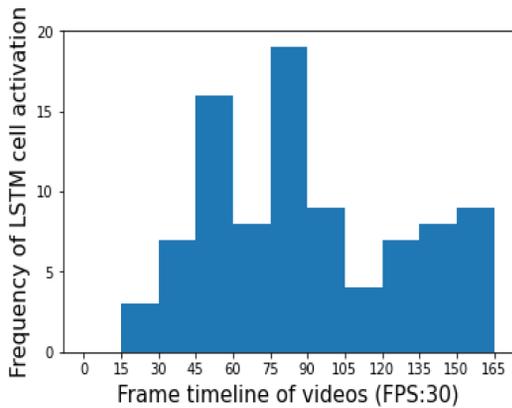


Fig. 3. LSTM cell activation results of truth data (Top) and deception data (Bottom). Frames per second (FPS) of video is 30.



**Fig. 4.** Histogram of LSTM cell activation frequency distribution table according to video timeline. Frames per second (FPS) of video is 30.

많은 데이터를 추가적으로 수집할 필요가 있다. 또한 얼굴표정 정보를 이용해 거짓말 여부를 탐지할 때 주목해야 할 중요 얼굴 위치 분석 모델의 개발이 필요하다.

결론적으로, 얼굴 특징점의 전반적인 변화를 이용한 비접촉식 거짓말 탐지 시스템 개발 및 유의시점 분석결과로부터 다음 결론을 얻을 수 있었다. 1) 본 연구를 통하여 얼굴 특징점의 전반적인 변화를 이용한 비접촉식 거짓말 탐지의 가능성을 확인할 수 있었다. 2) 또한 얼굴 특징점의 전반적인 변화를 이용하여 대답이 거짓말임을 판별하는데 유의미한 시점은 대답

**Table 1.** Frequency distribution table of the time when LSTM cell is activated for deception detection with test data

Timeline of videos	Frequency of LSTM cell activation
1~0.5 second before answer	0
0.5 second before answer~answer	3
During answer (0.5 second)	7
Answer end~0.5 second after answer	16
0.5~1 second after answer	8
1~1.5 second after answer	19
1.5~2 second after answer	9
2~2.5 second after answer	4
2.5~3 second after answer	7
3~3.5 second after answer	8
3.5~4 second after answer	9

Abbreviation; LSTM, Long short term memory.

직후와 대답 이후 1초부터 1.5초 이내이었다. 3) 그러므로 대답 직후부터 대답 이후 1.5초 이내의 얼굴 특징점의 전반적인 변화를 이용하여 거짓말과 진실을 판별할 가능성이 있음을 의미한다.

## 감사의 글

이 논문은 과학기술정보통신부의 재원으로 한국연구재단 바이오의료기술개발사업의 지원을 받아 수행된 연구입니다(NRF-2017M3A9E1064781).

## 참고문헌

1. P Ekman, and W V Friesen, (1971), Constants Across Cultures in the Face and Emotion. *JPSA*, 17(2): 124-129.
2. M David, and P Ekman, (2008), Facial expression analysis. *Scholarpedia*, 3(5): 4237.
3. PM Peter, R Boice, AL Fingeret, and WR Zwick, D Kolko, S Munroe and A Grunberger, (1984), Mid-level measurement of social anxiety in psychiatric and non-psychiatric samples. *Behaviour Research and Therapy* 22.6: 651-660.
4. M Yap, B Rajoub, H Ugail, R Zwiggelaar, (2011), Visual cues of facial behaviour in deception detection. *IEEE 2011 International Conference on Computer Applications and Industrial Electronics (ICCAIE)*, *IEEE*, 294-299.
5. M Owayjan, A Kashour, N Haddad, M Fadel and G Souki, (2012), The design and development of a lie detection system using facial micro-expressions. 2012 2nd international conference on advances in computational tools for engineering applications (ACTEA). *IEEE* 33-38.
6. Z Wu, B Singh, L Davis and V Subrahmanian, (2018), Deception detection in videos. *AAAI Conference on Artificial Intelligence*. 32.1: 1695-1702.
7. K Donald and P Shaw, (2015), *Fundamentals of Polygraph Practice*, 61-78, Academic Press.
8. B Adrian, and G Tzimiropoulos (2017), How far are we from solving the 2d & 3d face alignment

- problem? (and a dataset of 230,000 3d facial landmarks). *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. 1021-1030.
9. J Wan, Z Lai, L SHen, J Zhou, C Gao, G Xiao and Z Hou, (2021), Robust facial landmark detection by cross-order cross-semantic deep network. *Neural Networks* 136: 233-243.
10. H Palangi, L Deng, Y Shen, J Gao, X He, J Chen, X Song and R Ward, (2016), Deep sentence embedding using long short-term memory networks: Analysis and application to information retrieval. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 24.4: 694-707.

<저자정보>

남보름(대학원생), 김주영(대학원생),  
김명수(대학원생), 최진경(대학원생),  
조백환(부교수), 김인영(교수)